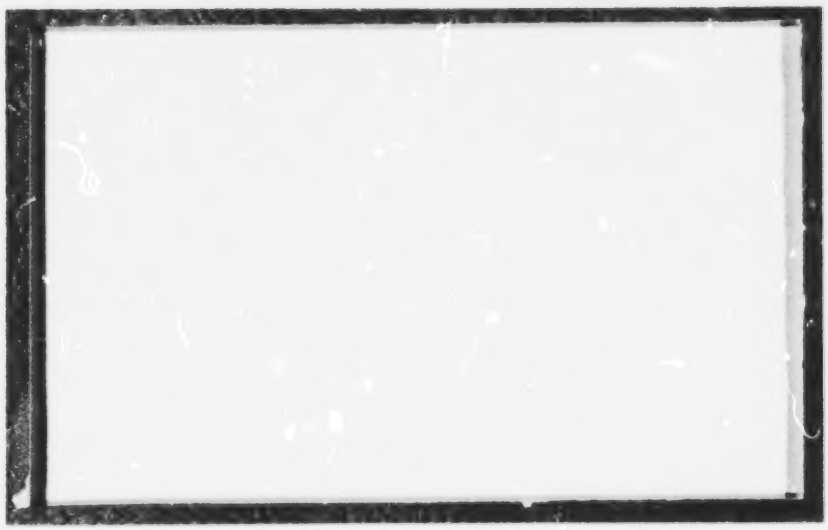
UNIVERSITÉ DE MONTRÉAL



CENTRE DE RECHERCHE SUR LES TRANSPORTS

CASE POSTALE 6128, SUCCURSALE A, MONTRÉAL, QUÉBEC, H3C 3J7

INFÉRENCE STATISTIQUE ET PROBLÈMES D'OPTIMISATION COMBINATOIRE DE GRANDE TAILLE

par

Christian Lardinois

Ce travail a été subventionné par le Centre de recherche et de développement de Transports Canada, par le programme F.C.A.C. du Ministère de l'éducation du Québec, ainsi que par le Prêt d'honneur Inc. de l'Université de Montréal.

Université de Montréal

Centre de recherche sur les transports - Publication # 145

Septembre 1979 - Révisé en mars 1980

RÉSUMÉ

L'estimation de l'optimum global des problèmes combinatoires de grande taille - estimation nécessaire à l'évaluation des solutions heuristiques - et l'interruption la plus opportune de l'exploration réalisée par les techniques de hill-climbing sont les deux questions traitées dans ce rapport. Quelques auteurs les ont abordées sous l'angle de l'inférence statistique en établissant une analogie avec la théorie des valeurs extrêmes. Ces travaux, qui fondent notre propre recherche, sont brièvement présentés et diverses critiques de fond sont formulées. Du point de vue de l'estimation de l'optimum global, un nouvel intervalle de confiance est défini, un exemple numérique relatif au problème d'affectation quadratique est présenté et nous proposons deux prolongements. En ce qui a trait à l'utilisation "économique" des procédures de hill-climbing, nous formulons une règle itérative ayant pour objet le choix rationnel de la taille des suites d'optima locaux qu'une telle procédure doit engendrer successivement jusqu'à ce que certaines conditions d'arrêt soient rencontrées.

1. INTRODUCTION

Pour certains problèmes d'optimisation combinatoire de grande taille, les méthodes d'exploration du type branch and bound s'avèrent souvent inefficaces (temps d'exécution trop long, place de mémoire requise trop importante). Ces problèmes "difficiles" sont nombreux: problème du chemin le plus long, problème du voyageur de commerce, problème du sac alpin, problème d'affectation quadratique, etc... Plusieurs parmi eux appartiennent à ce qu'on appelle la famille "NP-complete" et la plupart des auteurs (voir, entre autres, Garey et Johnson [2]) laissent entendre qu'aucun algorithme exact et efficace ne peut être inventé pour les résoudre. Pour solutionner ces problèmes "difficiles", on a recours à des algorithmes heuristiques.

L'objectif central de ce papier est l'évaluation de la qualité des solutions heuristiques, c'est-à-dire l'évaluation de l'écart entre la valeur de ces solutions et la valeur de la solution exacte. Dans ce but, c'est une estimation de la valeur de la solution exacte qui sera recherchée.

Il existe deux catégories principales de procédures heuristiques: les techniques séquentielles (constructive techniques), et les techniques d'exploration dirigée (hill-climbing techniques), ces dernières étant parfois couplées à des procédures de génération aléatoire (random sampling techniques) pour produire les solutions initiales qui leur sont nécessaires. Les méthodes séquentielles construisent une solution de manière itérative en prenant à chaque itération la "décision élémentaire" la plus favorable à l'optimisation. Les méthodes de hill-climbing démarrent avec une solution complète qu'elles améliorent progressivement par petits ajustements successifs choisis selon quelques règles systématiques, et cela jusqu'à ce qu'il ne soit plus possible d'améliorer la solution courante; la solution finale constitue un "optimum local" (Reiter et Sherman [13]); en appliquant une telle procédure à

différentes solutions initiales, on aboutit donc, en principe, à différents optima locaux parmi lesquels peut se trouver la solution exacte (l'optimum global).

Le problème de la proximité du meilleur optimum local atteint à un moment donné vis-à-vis de l'optimum global est lié au problème de choix qui consiste à décider s'il faut poursuivre la génération d'optima locaux, ou si au contraire il est opportun d'interrompre l'exploration. La formulation d'une règle d'arrêt constitue le second objectif de ce papier.

Les techniques de hill-climbing sont au coeur des considérations développées dans cet article. Non seulement très efficaces du point de vue des temps de calcul, ces techniques offrent la possibilité d'engendrer de nombreuses solutions localement optimales; cette possibilité est très avantageuse dans un contexte où il est souhaitable de disposer de plusieurs solutions alternatives (Ex.: un processus de planification où on veut évaluer plusieurs alternatives). Reiter et Sherman [13] ont montré comment toute procédure de ce type peut être interprétée moyennant l'introduction de ce qu'ils ont appelé une "structure de succession". Cette structure peut être comprise comme une partition de l'ensemble de toutes les solutions possibles en sous-ensembles disjoints, chacun de ceux-ci regroupant les solutions qui aboutiraient à un même optimum local si on les introduisait comme solutions initiales dans l'algorithme; cet optimum local est le successeur ultime commun aux éléments du sous-ensemble correspondant, et ce dernier a la forme d'un arbre. A cette structure de succession, on peut associer une "structure de probabilité": si la solution initiale est choisie au hasard, c'est-à-dire si elle provient d'une procédure de génération aléatoire, la probabilité de sélectionner une solution initiale dont le successeur ultime soit l'optimum local $\bar{\rho}$ est donnée par [1]

$$\text{prob}(\bar{\rho}) = \frac{|T(\bar{\rho})|}{|M|}$$

où M est l'ensemble de toutes les solutions possibles et $T(\bar{\rho})$ représente le sous-ensemble de M qui regroupe toutes les solutions qui aboutiraient à $\bar{\rho}$ si on leur appliquait l'algorithme. En d'autres termes, la combinaison d'une procédure de

1. Si A est un ensemble, alors |A| représente le nombre cardinal de A.

hill-climbing et d'une procédure de génération aléatoire induit une distribution de probabilité (définie par la relation ci-dessus) sur le sous-ensemble des optima locaux associés à l'algorithme de hill-climbing envisagé.

La suite du papier est organisée comme suit: la section 2 précise les notations et introduit l'inférence statistique comme cadre théorique; la section 3 présente d'abord l'essentiel des recherches qui fondent notre étude en ce qui a trait à l'estimation de l'optimum global; diverses critiques sont formulées et un nouvel intervalle de confiance est défini; un exemple numérique est présenté et deux prolongements sont proposés; la règle d'arrêt fait l'objet de la section 4 et la dernière section conclut ce papier.

2. L'APPROCHE CHOISIE

L'approche adoptée ici est suffisamment générale pour être appliquée à de nombreux problèmes d'optimisation combinatoire. L'ensemble de toutes les solutions possibles d'un tel problème et la fonction objectif peuvent être représentés par M et X. A toute solution $\rho \in M$, la fonction X définie sur M associe une valeur $X(\rho)$. Pour fixer les idées dans la suite, on supposera avoir affaire à un problème d'affectation quadratique [10]: on dispose de n activités à affecter à n localisations; il existe n! solutions possibles, c'est-à-dire n! applications biunivoques possibles ρ de l'ensemble A des activités sur l'ensemble $\mathcal{L}$ des localisations; chaque application possible $\rho: A \twoheadrightarrow \mathcal{L}$ représente une permutation particulière des n éléments de A. A toute permutation ρ est associé un coût $X(\rho)$, et le problème d'optimisation consiste à trouver une permutation $\rho^* \in M$ de coût minimal, c'est-à-dire telle que

$$X(\rho^*) \leq X(\rho) \qquad \text{pour tout } \rho \in M$$

Les variables seront représentées par des lettres majuscules; les valeurs prises ou

réalisables seront représentées par les lettres minuscules correspondantes: $x_i \equiv X(\rho_i)$, $x^* \equiv X(\rho^*)$, etc...

Contrairement à de nombreuses recherches (theory of computational complexity; voir notamment Garey et Johnson [2]), notre approche a pour but non pas l'évaluation, a priori, des performances minimales d'un algorithme heuristique indépendamment des données spécifiques caractérisant chaque problème, mais bien l'évaluation de la qualité des solutions effectivement engendrées: pour un problème donné, on veut estimer l'écart entre les solutions obtenues et l'optimum global en produisant une estimation de celui-ci.

Si l'on pouvait énumérer explicitement toutes les n! permutations possibles, on serait à même de calculer la probabilité qu'un élément ρ_i choisi au hasard dans M ait une valeur inférieure ou égale à n'importe quelle grandeur x fixée a priori

$$\text{prob}\,[X(\rho_i) \leq x] = \frac{|\{\rho : \rho \in M; X(\rho) \leq x\}|}{|M|}$$

Ces probabilités fourniraient la fonction de répartition des images x_i des éléments $\rho_i \in M$. En d'autres termes, dès qu'on considère une procédure de génération aléatoire des éléments $\rho_i \in M$, la fonction objectif X devient une variable statistique aléatoire à laquelle est associée une distribution de probabilité caractérisant la population-mère M. Cette distribution des valeurs $X(\rho_i) \equiv x_i$ où i=1,2,...,n! est malheureusement inconnue; tout ce qu'on sait, c'est qu'elle est bornée à gauche par l'optimum global que nous cherchons.

La première idée qui vient naturellement à l'esprit est de faire appel à l'inférence statistique en étudiant la distribution d'un échantillon aléatoire tiré de la population-mère à l'aide d'une procédure de génération aléatoire. L'inférence statistique classique (paramétrique) n'est cependant capable d'associer une mesure de confiance aux conclusions qu'elle propose quant à la distribution-mère que si elle dispose d'un minimum d'information sur cette population-mère ou si elle propose une hypothèse à son sujet. Formuler une hypothèse sur la distribution de la population-mère paraît cepen-

dant très hasardeux: a priori, cette population discrète ne répond à aucune loi, à aucune structure particulière. Pour échapper à la nécessité de formuler une hypothèse sur la distribution-mère, il faut se situer dans le cadre de l'inférence statistique non-paramétrique (ou méthodes de libre distribution [3]), laquelle a précisément pour caractéristique essentielle de rechercher des informations indépendantes de la distribution-mère.

3. ESTIMATION DE L'OPTIMUM GLOBAL

3.1. Analogie avec la théorie des valeurs extrêmes

Lorsqu'on utilise une procédure de hill-climbing, il est facile d'engendrer de nombreuses solutions heuristiques. Si ρ_1^h, ρ_2^h, ..., ρ_N^h sont N permutations construites à l'aide d'une telle procédure, ces N solutions, en tant qu'optima locaux, contiennent nécessairement une certaine information. L'approche décrite dans la suite de cet article s'efforce de tirer parti de toute cette information. (Lorsque nous parlerons d'heuristiques, il ne s'agira donc plus que des techniques du type hill-climbing. Les solutions initiales nécessaires à ces techniques seront engendrées par l'une ou l'autre procédure de génération aléatoire [8]).

L'idée centrale (Reiter et Sherman [13]) est de proposer une hypothèse non pas sur la distribution de la population-mère, mais plutôt sur la distribution des optima locaux (cf. introduction). Cette idée est au centre des travaux de McRoberts [12]; l'approche imaginée par celui-ci s'inspire de la théorie des valeurs extrêmes (Gumbel [6]), et plus particulièrement d'un théorème énoncé pour la première fois en 1928 par Fisher et Tippett [1]:

> Lorsque la taille m de chacun des N échantillons indépendants issus d'une distribution quelconque f(x) bornée à gauche devient très grande, la distribution des valeurs minimales $x_i (i=1,\ldots,N)$ de ces N échantillons approche une distribution limite connue, indépendante de la distribution-mère f(x), et dite "distribution de Weibull"[14].

La fonction de répartition et la densité de probabilité de cette distribution asymptotique sont respectivement:

$$\Phi(x) = 1 - \exp\left[- \left(\frac{x-a}{b}\right)^{c} \right]$$

$$\delta(x) = \frac{c}{b} \left(\frac{x-a}{b}\right)^{c-1} \exp\left[-\left(\frac{x-a}{b}\right)^{c} \right] \qquad x \geq a \geq 0;\ c>0;\ b>0$$

Les paramètres a, b et c sont les paramètres de localisation, d'échelle et de forme. Le premier paramètre (a), borne inférieure de la distribution en question, est en même temps la borne inférieure de la distribution-mère. L'idée de McRoberts est la suivante. Chaque succession de solutions construites durant l'application du hill-climbing à une solution initiale (aléatoire) jusqu'à un optimum local x_i^h est interprétée comme l'un des échantillons de taille m du théorème de Fisher-Tippett; l'optimum local x_i^h en est la valeur minimale. Avec N optima locaux x_i^h (i=1,...,N), on se trouve donc dans une situation analogue à celle décrite dans le théorème en question et il semble logique d'imaginer que la distribution des optima approche la loi Weibull. L'estimation de son paramètre de localisation (a) fournira une approximation de l'optimum global.

Golden et Alt [4,5] ont repris cette approche en modifiant légèrement l'argumentation: pour eux, ce n'est plus la succession des solutions construites durant l'application du hill-climbing à une solution initiale (aléatoire) qui constitue l'un des N échantillons du théorème de Fisher-Tippett; chacun de ces échantillons est plutôt formé du sous-ensemble particulier et implicite de solutions possibles qui aboutiraient au

même optimum local si on leur appliquait le hill-climbing (cf. introduction). Mis à part cet "ajustement", le principe de la méthode est inchangé: une estimation de l'optimum global est obtenue par estimation du paramètre de localisation de la loi Weibull à laquelle on identifie, par hypothèse, la distribution des optima locaux résultant de l'application du hill-climbing à des solutions initiales aléatoires.

En fait, l'estimation ponctuelle du paramètre de localisation à partir d'un échantillon de taille N se réalise en même temps que celle des deux autres paramètres b et c. La littérature propose différentes méthodes. La procédure de Harter-Moore [7] qui est basée sur le principe du maximum de vraisemblance est recommandée par différents auteurs (Mann *et al.* [11], Zanakis [15]).

L'obtention d'un intervalle de confiance est un problème beaucoup plus compliqué que celui de l'estimation ponctuelle. L'intervalle défini par Golden et Alt [5] exploite une propriété à laquelle la loi Weibull obéit toujours, à savoir: $\Phi(a+b)=1-e^{-1}$; si $x_i^h \equiv X(\rho_i^h)$, $(i=1,\ldots,N)$ forment un échantillon de N optima locaux, on a [1]

$$\text{prob}\{x_{(1)}^h \le a+b\} = 1 - \text{prob}\{x_{(1)}^h > a+b\}$$

$$= 1 - \{1-\Phi_{X_1^h}(a+b)\}\{1-\Phi_{X_2^h}(a+b)\}\ldots\{1-\Phi_{X_N^h}(a+b)\}$$

$$= 1 - (e^{-1})^N = 1 - e^{-N}$$

En utilisant l'estimation $\hat{b}$ du paramètre d'échelle, on obtient ainsi un intervalle de confiance approximatif à $100\,(1-e^{-N})\%$

$$\text{prob}\{x_{(1)}^h - \hat{b} \le x^* \le x_{(1)}^h\} \simeq 1 - e^{-N}$$

---

1. $x_{(1)}^h \le x_{(2)}^h \le \ldots \le x_{(N)}^h$ sont les observations $x_1^h, x_2^h, \ldots, x_N^h$ ordonnées par ordre de valeurs croissantes.

Pour fonder davantage leur approche, Golden et Alt ont soumis l'hypothèse Weibull au test de Kolmogorov-Smirnov (Gibbons [3]). Ce test d'ajustement compare la distribution hypothétique avec la distribution empirique. Quant à l'hypothèse d'indépendance des N observations à partir desquelles est réalisée l'estimation ponctuelle des trois paramètres, Golden et Alt se sont servis de tests qui vérifient le caractère aléatoire de suites d'observations (tests based on runs [3]). Tous ces tests ont conduit chaque fois à l'acceptation des hypothèses en question.

Malgré la modification apportée par Golden et Alt à l'argumentation de McRoberts et malgré le succès des tests statistiques, il reste que

- l'indépendance des observations est sujette à caution puisque, en tant qu'optima locaux, elles constituent les sommets d'arbres disjoints formant la structure de succession qui est déterminée par l'heuristique sur l'ensemble M de toutes les solutions possibles; comme points particuliers d'une structure associée à une heuristique particulière, ces observations peuvent difficilement être qualifiées d'indépendantes;
- l'indépendance qui théoriquement doit caractériser les éléments formant, par analogie, chacun des N échantillons du théorème de Fisher-Tippett est également discutable puisque ces éléments partagent la propriété d'aboutir au même optimum local si on leur appliquait le hill-climbing.

Enfin, il faut souligner que le théorème de Fisher-Tippett suppose qu'on ait affaire à des distributions continues, ce qui n'est pas le cas avec les problèmes combinatoires que nous envisageons ici: la pertinence même de l'approche se trouve donc hypothéquée par l'approximation implicite de distributions discrètes par des distributions continues. Toutefois, en l'absence d'une théorie précisant les conditions dans lesquelles un théorème formulé pour des variables continues peut être étendu ou ajusté au cas des populations discrètes, nous supposerons que le phénomène d'explosion combinatoire autorise l'approximation en question, à condition cependant que soit introduite la distinction présentée ci-dessous.

La distribution empirique (sample or empirical cumulative distribution function) des N optima locaux x_1^h, x_2^h, ..., x_N^h est bien une approximation de la distribution de probabilité induite par l'association du hill-climbing et d'une procédure de génération aléatoire des solutions initiales — distribution définie par Reiter et Sherman (cf. introduction) —, mais on ne peut faire l'hypothèse que cette distribution induite approche la loi Weibull car les conditions du théorème de Fisher-Tippett ne sont pas satisfaites. Chaque fois que l'un des optima se répète plusieurs fois dans l'échantillon de taille N, cela revient en effet à utiliser plusieurs fois le même échantillon implicite. Dans nos premiers essais numériques sur le problème d'affectation quadratique, nous nous sommes ainsi retrouvés dans un cas [1] où, parmi les N=100 optima de l'échantillon destiné à estimer la loi Weibull, le plus petit apparaissait 24 fois; c'est au droit de ce point que s'est situé l'écart maximum entre la distribution empirique et la courbe Weibull hypothétique; le test de Kolmogorov-Smirnov a conduit au rejet de l'hypothèse d'ajustement. Par contre, si dans l'échantillon de taille N on ne retient que les $R \leq N$ optima distincts x_1^{hd}, x_2^{hd}, ..., x_R^{hd}, on supprime du même coup les répétitions des mêmes échantillons implicites; on dispose ainsi de R valeurs extrêmes "extraites" de R échantillons distincts "issus"de la même population-mère, R extrema à propos desquels l'hypothèse Weibull semble cette fois justifiée. En fait, en ne conservant que les R optima distincts, tout se passe comme si cet échantillon de taille R était tiré d'une pseudo-population F qui se comporte comme une population continue. En résumé, ce n'est pas à propos de la *distribution induite* dont parlaient Reiter et Sherman que l'on peut faire l'hypothèse Weibull, mais bien à propos de la distribution de cette pseudo-population F; c'est cette façon de procéder qui est adoptée dans la suite du texte; la distribution en question sera dite *pseudo-distribution* des optima locaux.

---

1. Il s'agissait d'un problème de taille 11 : 11 activités à affecter à 11 localisations.

La distinction introduite ci-dessus résulte du caractère discret des problèmes traités; elle devient dérisoire pour les problèmes de grande taille avec lesquels les N optima initialement engendrés sont presque tous distincts; lorsque la taille des problèmes augmente, le phénomène d'explosion combinatoire nous déplace du côté des situations où l'on aurait affaire à des distributions de type continu, des situations où l'on peut considérer que l'estimation de la "pseudo-distribution" des optima locaux est aussi une estimation de la distribution "induite" (et où l'analogie avec Fisher-Tippett est d'autant plus fondée). Il reste bien sûr la question de savoir à partir de quelle taille minimale des problèmes on peut admettre cette équivalence. Cette question n'est pas purement académique: elle est d'une importance capitale pour ce qui est de la règle d'arrêt qui fait l'objet de la section 4, règle pour laquelle c'est précisément la distribution "induite" qui joue un rôle central.

Remarquons que la distinction entre la distribution induite et la pseudo-distribution des optima locaux n'élimine pas pour autant les objections mentionnées plus haut quant à l'indépendance des observations ainsi que celle des éléments formant les échantillons implicites: ces objections subsistent pour les R optima distincts et les échantillons dont ils sont "issus". Toutefois, jusqu'à preuve du contraire, nous supposerons qu'elles n'invalident pas l'analogie avec le théorème de Fisher-Tippett.

3.2. Un nouvel intervalle de confiance

Soient x_i^{hd} (i=1,...,R) R optima locaux distincts. Avec la distinction introduite ci-dessus, l'intervalle de confiance défini par Golden et Alt devient:

$$\text{prob}\{x_{(1)}^{hd} - \hat{b} \leq x^* \leq x_{(1)}^{hd}\} \simeq 1 - e^{-R}$$

où $x_{(1)}^{hd} \leq x_{(2)}^{hd} \leq \ldots \leq x_{(R)}^{hd}$ sont les observations $x_1^{hd}, x_2^{hd}, \ldots, x_R^{hd}$ ordonnées par ordre de valeurs croissantes, et où $\hat{b}$ provient de l'estimation de la "pseudo-

distribution", estimation qui est réalisée à partir de ces R observations. Cet intervalle apporte en fait une information relativement faible. En première approximation, en effet, sa largeur $\hat{b}$ est mesurée par [1]

$$\hat{b} \simeq x^{hd}_{([0.63R]+1)} - \hat{a}$$

où [0.63R] est le plus grand entier inférieur à 0.63R; cette largeur est donc toujours très grande par rapport à l'étendue de variation $x^{hd}_{(R)} - x^{hd}_{(1)}$ des R observations. Par exemple, dans le cas d'un problème d'affectation quadratique de taille 24 (Lardinois [9]), un échantillon de R=50 optima distincts a fourni un intervalle dont la largeur représentait 61% de l'étendue des observations (degré de confiance quasi égal à 100%). La faiblesse de cet intervalle résulte de l'insensibilité du degré de confiance qui lui est associé par rapport à ses limites supérieures et inférieures: même pour des valeurs relativement petites de R, on a toujours $1-e^{R} \simeq 1$.

En suivant sensiblement le même raisonnement que Golden et Alt, il est heureusement possible d'aboutir à un intervalle de confiance qui soit vraiment significatif. Soit S un nombre réel quelconque. La fonction de répartition de la loi Weibull obéit toujours à la relation suivante:

$$\Phi(a + \frac{b}{S}) = 1 - \exp\left[-\left(\frac{1}{S}\right)^{c}\right]$$

Les optima locaux distincts x^{hd}_{i} (i=1,...,R) d'un échantillon de taille R doivent donc satisfaire

$$\text{prob}\{x^{hd}_{(1)} \leq a+\frac{b}{S}\} = 1 - \text{prob}\{x^{hd}_{(1)} > a+\frac{b}{S}\}$$

$$= 1 - \{1 - \Phi_{x^{hd}_1}(a+\frac{b}{S})\}\{1 - \Phi_{x^{hd}_2}(a+\frac{b}{S})\}\ldots\{1 - \Phi_{x^{hd}_R}(a+\frac{b}{S})\}$$

$$= 1 - \exp(-R/S^{c})$$

---

1. Cette valeur approchée pour b résulte de la propriété $\Phi(a+b)=1-e^{-1} \simeq 0.63$

En introduisant les estimations $\hat{b}$ et $\hat{c}$ des paramètres d'échelle et de forme, on obtient un intervalle (approximatif) dont la largeur et le degré de confiance sont cette fois explicitement reliés l'un à l'autre

$$\text{prob } \{x^{hd}_{(1)} - \frac{\hat{b}}{S} \leq x^* \leq x^{hd}_{(1)}\} \simeq 1 - \exp\ (-R/S^{\hat{c}})$$

Si on pose $1 - \exp\ (-R/S^{\hat{c}}) = 1 - \alpha$, la relation qui relie le degré de confiance au nombre S, au paramètre de forme c et au nombre R d'observations s'écrit

$$S = (-\frac{R}{\ell n \alpha})^{1/\hat{c}}$$

Pour des valeurs fixées de R, $\hat{b}$ et $\hat{c}$, une augmentation du degré de confiance $1-\alpha$ est suivie à juste titre d'une augmentation de la largeur de l'intervalle.
On voit aussi que cette fois, le degré de confiance, et donc la largeur de l'intervalle, sont fonction du paramètre de forme: pour des valeurs de R et α fixées, des distributions de même $\hat{b}$ mais dont les valeurs de $\hat{c}$ sont croissantes, conduisent à des intervalles plus larges. Ceci traduit bien le fait qu'à un paramètre de forme de plus grande valeur correspond un étalement plus prononcé de la partie gauche de la courbe Weibull.

Les résultats présentés ci-dessous sont relatifs à un problème d'affectation quadratique de taille 48. Les données de ce problème ont été engendrées de façon aléatoire en utilisant le générateur de nombres aléatoires (RANF) du Centre de calcul de l'Université de Montréal. Ce même générateur a également servi à engendrer les permutations aléatoires initiales à l'aide d'un algorithme[1] dû à Moses et Oakford. Les optima locaux ont été construits avec l'algorithme[2] ACTIV de Los [10]. L'initialisation de la procédure de Harter-Moore [7] a été réalisée avec les expressions utilisées par Golden et Alt [5]. Les résultats furent les suivants

---

1. Shuffling technique - Knuth [8] p. 125.
2. Steepest-descent pairwise interchange procedure.

(N=R=140 optima locaux distincts):

$$x^{hd}_{(1)} = 529990157; \quad x^{hd}_{(140)} = 537117008$$

$$\hat{a} = 529616961; \quad \hat{b} = 3761353; \quad \hat{c} = 2.9105$$

L'intervalle de confiance à 95% est

$$\text{prob}\ (528986249 \leq x^* \leq 529990157) \simeq 0.95$$

La largeur $\hat{b}/S$ de cet intervalle représente 14% de l'étendue de variation $x^{hd}_{(140)} - x^{hd}_{(1)}$ des 140 observations. Sa largeur relative $(\hat{b}/S)100/x^{hd}_{(1)}$ est de 0.19%.

La valeur observée de la statistique D_R du rest de Kolmogorov-Smirnov est 0.0357. La valeur critique correspondante (Gibbons [3]) étant 0.1148 au niveau de signification 0.05, l'hypothèse Weibull n'est pas rejetée. La figure 1 illustre cet ajustement: la distribution empirique des 140 observations est symbolisée par les croix et la ligne continue représente la courbe Weibull théorique correspondant aux valeurs estimées des 3 paramètres.[1]

3.3. Combinaison d'heuristiques et extrema d'échantillons d'optima

Les $R \leq N$ observations distinctes dont on identifie la distribution avec la loi Weibull ne sont pas indépendantes: la "structure de succession" propre à l'algorithme utilisé a pour conséquence que ce sont telles solutions et non telles autres qui deviennent des optima locaux; en d'autres mots, les R observations sont "heuristiquement" liées les unesaux autres. En fait, tout ce que requiert l'analogie avec le théorème de Fisher-Tippett, c'est que les observations puissent être interprétées comme les valeurs extrêmes d'échantillons indépendants "issus" d'une même population. En conséquence, au lieu de n'utiliser qu'une heuristique pour engendrer les R extrema, c'est au contraire plusieurs heuristiques qui peuvent intervenir. L'utilisation de différents algorithmes aurait l'avantage

---

1. L'origine des axes a été translatée à l'origine de la courbe théorique. Il faut donc lire tous les nombres en abscisse en leur ajoutant la valeur du paramètre a.

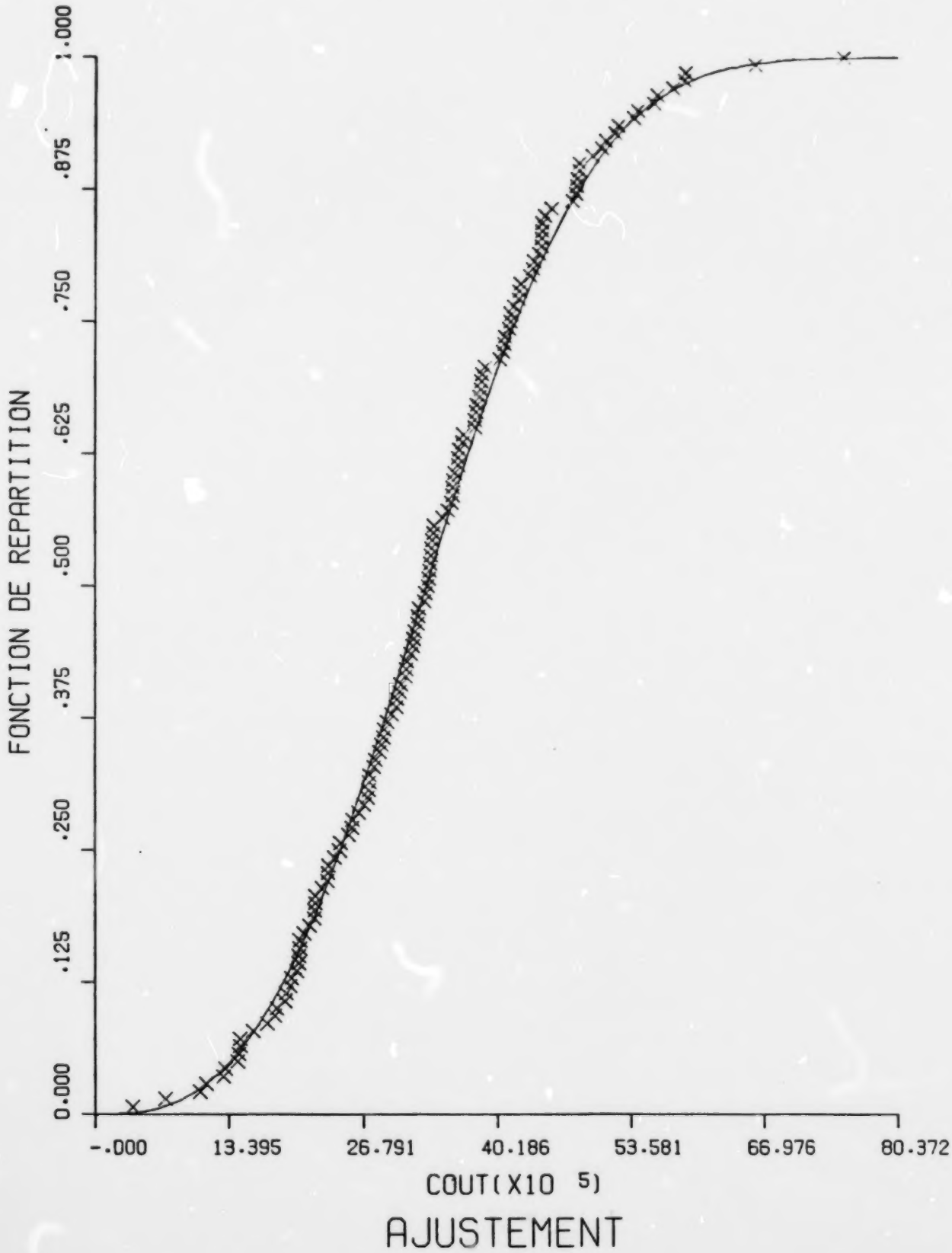

FONCTION DE REPARTITION
0.000
.125
.250
.375
.500
.625
.750
.875
1.000
-.000
13.395
26.791
40.186
53.581
66.976
80.372
COUT(X10 5)
AJUSTEMENT

Figure 1

- de produire des observations parmi lesquelles certaines ne seraient plus "heuristiquement" dépendantes les unes des autres
- de multiplier le nombre d'observations pouvant jouer un rôle puisque l'ensemble des optima locaux susceptibles d'apparaître serait cette fois consitué de la réunion des ensembles d'optima locaux respectivement liés aux différentes heuristiques.

Nous avons fait remarquer également que les éléments formant chacun des échantillons implicites ne sont pas indépendants: ils partagent la propriété d'aboutir au même optimum local. Cette constatation suggère un autre prolongement. L'idée est de réaliser explicitement la production d'échantillons d'optima locaux (résultant à nouveau d'une procédure de génération aléatoire suivie d'un hill-climbing) et dans chacun d'eux d'y extraire la valeur minimale. Soient donc E échantillons G_i de m optima locaux

$$G_i = \{x_1^{h,i}, x_2^{h,i}, \ldots, x_m^{h,i}\} \qquad (i=1,\ldots,E)$$

Ensuite, dans chacun des E ensembles G_i, on prend la valeur minimale $x_{(1)}^{h,i}$; on aboutit ainsi à un "super échantillon"

$$x_{(1)}^{h,1}, x_{(1)}^{h,2}, \ldots, x_{(1)}^{h,E} \qquad \text{avec } x_{(1)}^{h,i} = \text{Min } \{x_j^{h,i}, j=1,\ldots,m\}$$

dont chaque élément est la valeur extrême d'un échantillon de taille m issu d'une même population, à savoir la population dont la distribution est la distribution "induite" définie précédemment. On se trouve alors une nouvelle fois dans une situation analogue à celle du théorème de Fisher-Tippett: la distribution de ces E extrema $x_{(1)}^{h,i}$ (i=1,...,E) doit approcher la loi Weibull. [1] L'estimation de ses trois paramètres fournira une nouvelle estimation ponctuelle $\hat{x}^*=\hat{a}$ et un nouvel intervalle de confiance pour l'optimum global

---

1. Ici, rien n'interdit que dans un même ensemble G_i on retrouve plusieurs fois la même observation. Il se peut aussi que le même extremum $x_{(1)}^{h,i}$ se retrouve plusieurs fois dans le "super-échantillon".

$$\text{prob}\ \{x^{h,(1)}_{(1)} - \frac{\hat{b}}{S} \leq x^* \leq x^{h,(1)}_{(1)}\} \simeq 1 - \alpha$$

où $$x^{h,(1)}_{(1)} = \text{Min}\ \{x^{h,i}_{(1)},\ i=1,\ldots,E\}$$

$$S = (-\frac{E}{\ell n \alpha})^{1/\hat{c}}$$

Soulignons que précédemment, chaque échantillon dont on tirait la valeur extrême était implicite, ce qui n'est plus le cas ici. D'autre part, chacun de ces échantillons était "issu" de la population-mère M; cette fois, étant constitués d'optima locaux effectivement engendrés, les échantillons sont issus de la population dont la distribution de probabilité est celle qui caractérise les optima susceptibles de résulter de l'association hill-climbing/génération aléatoire, à savoir la distribution "induite" mentionnée plus haut. Remarquons aussi qu'à présent, l'indépendance des m observations constituant chacun des E échantillons G_i peut être testée. Il va sans dire qu'un test d'ajustement comme celui de Kolmogorov-Smirnov est également requis puisque nous continuons à utiliser dans un cas discret un théorème appartenant à une théorie formulée pour des distributions continues.

4. REGLE D'ARRET

Revenons à l'estimation de l'optimum global à partir des $R \leq N$ optima locaux distincts engendrés par un algorithme de hill-climbing. Après avoir procédé à cette estimation, le problème qui se pose est de savoir s'il est opportun de poursuivre l'exploration de l'espace des solution possibles, ou au contraire s'il est plus judicieux de s'arrêter. Il s'agit là du problème de l'utilisation "économique" des techniques de hill-climbing.

Si $x^h \equiv X(\rho^h)$ est le meilleur des optima engendrés, l'estimation $\hat{x}^* = \hat{a}$ de l'optimum global fournit directement une approximation du *potentiel d'amélioration* $PA(\rho^h)$ par

rapport à cette meilleure permutation ρ^h

$$PA(\rho^h) = X(\rho^h) - x^* \simeq x^h - \hat{x}^*$$

Si C_o représente l'utilité unitaire, c'est-à-dire le bénéfice que constitue la réduction d'une unité de la valeur $X(\rho)$ d'une permutation ρ, alors la grandeur $PA(\rho^h).C_o$ représente l'utilité du gain maximum possible; si C_S représente le coût de calcul d'une exploration supplémentaire, la règle suivante est une première façon de décider de l'arrêt ou de la continuation de l'exploration

$$\text{si } PA(\rho^h).C_o > C_S \Rightarrow \text{poursuivre l'exploration}$$
$$\text{si } PA(\rho^h).C_o \leq C_S \Rightarrow \text{stop}$$

La valeur du paramètre C_S est fonction de l'algorithme de hill-climbing utilisé, de la taille du problème et bien sûr de l'ordinateur disponible. Le choix d'une valeur pour le paramètre C_o dépend de la signification des valeurs de la fonction objectif; quand ces valeurs et les coûts de calcul sont exprimés en unités monétaires, on a $C_o = 1$; quand les valeurs de la fonction objectif n'ont pas une signification monétaire, il doit exister un moyen d'évaluer chaque unité gagnée sur la valeur de la fonction objectif; si par exemple on évalue à \$300 la dépense autorisée pour l'acquisition de 10% du potentiel d'amélioration $PA(\rho^h)$, on a $C_o = 300/[0.10PA(\rho^h)]$ exprimé en \$/unité.

4.1. La probabilité d'amélioration et la valeur espérée du gain

La règle d'arrêt formulée ci-dessus est trop élémentaire: même si le potentiel d'amélioration est important, la probabilité qu'une exploration additionnelle aboutisse à une permutation meilleure que ρ^h peut être très faible.

Ce qui nous intéresse, c'est la loi décrivant la probabilité d'engendrer, à partir

d'une solution initiale aléatoire, un optimum local meilleur qu'une valeur donnée x^h. Ce qui nous intéresse, c'est la distribution définie par Reiter et Sherman, à savoir la distribution "induite" décrivant le comportement du hill-climbing et de la procédure de génération aléatoire des solutions initiales (voir 1 et 3.1). Dans la mesure où l'on peut considérer que la "pseudo-distribution" (Weibull) des optima locaux estimée à partir des R optima distincts de l'échantillon original de taille N est simultanément une estimation de la distribution "induite" en question [1], c'est-à-dire une estimation de la loi de probabilité décrivant la distribution des optima susceptibles d'être engendrés, la *probabilité d'amélioration* qui nous intéresse est fournie par

$$\text{prob}\ (X \leq x^h) = \hat{\Phi}(x^h)$$

où $\hat{\Phi}(x)$ est la fonction de répartition de la loi Weibull ainsi estimée.

Parallèlement au potentiel et à la probabilité d'amélioration, un autre critère (Reiter et Sherman [13]) doit jouer un rôle: il s'agit de la *valeur espérée du gain* fournie par une observation additionnelle. McRoberts [12] calcule cette grandeur de la façon suivante [2]

$$E(\text{Gain}) \simeq \int_{\check{x}^*}^{x^h} (x^h - x)\ \hat{\delta}(x)\ dx$$

où $\hat{\delta}(x)$ est la densité de la loi Weibull estimée.

4.2. Elaboration d'une règle générale intégrant les trois critères

La probabilité d'amélioration et la valeur espérée du gain définies ci-dessus sont relatives à une exploration additionnelle. En fait, l'obtention d'un optimum local

---

1. Cette équivalence n'est justifiée que pour les problèmes de grande taille (voir 3.1). Dans la suite du texte, nous supposerons que cette équivalence est permise.
2. Rappelons cependant que McRoberts ne faisait pas de distinction entre la distribution "induite" et la "pseudo-distribution" des optima locaux.

meilleur que x^h peut nécessiter un nombre non négligeable d'observations additionnelles. Il serait donc judicieux que l'évaluation des deux critères en question soit réalisée non pas vis-à-vis d'une observation additionnelle mais plutôt vis-à-vis de plusieurs observations additionnelles. D'autre part, la valeur espérée du gain n'a vraiment de signification que relativement au gain maximum possible, à savoir le potentiel d'amélioration défini plus haut qui constitue un troisième critère. Il serait donc également intéressant qu'une relation de comparaison entre la valeur espérée du gain et le potentiel d'amélioration $PA(x^h)$ intervienne.

La suite de ce texte propose une règle itérative définissant

- le nombre d'optima locaux qu'il faut engendrer à chaque itération pour contrôler simultanément la probabilité d'amélioration sur chacune de ces suites d'observations;
- des conditions d'arrêt basées sur la comparaison de la valeur espérée du gain sur la suite d'observations de l'itération subséquente avec le potentiel d'amélioration et avec le coût total de calcul de cette itération subséquente.

Calcul de la probabilité d'amélioration sur K observations additionnelles

La probabilité d'amélioration sur une "épreuve" (une exploration) additionnelle étant fournie par $\hat{\Phi}(x^h)$, dans K épreuves la densité de probabilité de la variable aléatoire Z (Z = nombre de fois qu'une amélioration apparaît dans K épreuves) est fournie par la loi binomiale

$$f(z) = \binom{K}{z} [\hat{\Phi}(x^h)]^z [1 - \hat{\Phi}(x^h)]^{K-z} \qquad z=0,1,2,\ldots,K$$

Il résulte que la probabilité d'amélioration sur K observations est donnée par la probabilité d'avoir au moins une des K observations $\leq x^h$

$$1 - \text{prob}\,(Z=0) = 1 - [1 - \hat{\Phi}(x^h)]^K$$

Si on fixe cette probabilité à une valeur α, le nombre d'observations additionnelles qui sont nécessaires pour garantir ce niveau de probabilité α doit être au moins égal

à la valeur de K obtenue en résolvant l'équation

$$1 - [1 - \hat{\Phi}(x^h)]^K = \alpha$$

ou $$K = \frac{\log (1-\alpha)}{\log[1 - \hat{\Phi}(x^h)]}$$

Calcul de la valeur espérée du gain sur K observations additonnelles

Cette grandeur est fournie par la valeur espérée de la variable aléatoire $x^h - X_{(1)}$ où $X_{(1)}$ représente la plus petite valeur observée dans un échantillon de taille K issu de la population dont la distribution n'est rien d'autre que la distribution "induite" dont il fut question précédemment, c'est-à-dire une population de densité $\delta(x)$ (Weibull). La densité de probabilité de la variable $X_{(1)}$ est donc la densité marginale de la plus petite variable statistique ordonnée correspondant à un échantillon de taille K issu de $\delta(x)$ (Gibbons [3])

$$p_{X_{(1)}}(x) \simeq K [1 - \hat{\Phi}(x)]^{K-1} \hat{\delta}(x) \quad .$$

D'où la valeur espérée du gain sur K observations

$$E_K(\text{Gain}) \simeq K \int_{\hat{x}^*}^{x^h} (x^h - x) [1 - \hat{\Phi}(x)]^{K-1} \hat{\delta}(x) \, dx$$

L'expression proposée par McRoberts est un cas particulier (K=1) de la relation définie ci-dessus.

Description de la procédure

1. Engendrer N_1 optima locaux[1] et poser i=1.
2. Itération i
   - Soit R_i le nombre d'optima distincts parmi les N_i optima qu'on a construits[2] et soit $x^{i,h}$ la valeur du meilleur d'entre eux[3].

---

1. N_1 doit évidemment être suffisant pour que l'estimation réalisée durant la première itération soit pertinente.
2. Rappelons qu'on se situe ici dans le cas des problèmes de grande taille avec lesquels $R_i \simeq N_i$.
3. Pour i>1, on a $x^{i,h} \leq x^{i-1,h}$.

- Estimer les trois paramètres de la loi Weibull à laquelle on identifie la distribution des R_i optima $\Rightarrow \hat{a}_i$, $\hat{b}_i$, $\hat{c}_i$.
- Choisir le niveau de probabilité d'amélioration que l'on veut garantir sur les N_{i+1} observations éventuelles de l'itération i+1; soit α_i ce niveau.
- Calculer N_{i+1} en résolvant l'équation[1]

$$N_{i+1} = \text{PPESE} \frac{\log(1-\alpha_i)}{\log[1-\hat{\Phi}_i(x^{i,h})]}$$

où $\hat{\Phi}_i(x)$ est la loi Weibull avec les valeurs $\hat{a}_i$, $\hat{b}_i$ et $\hat{c}_i$ pour les paramètres.

- Calculer la valeur espérée du gain sur les N_{i+1} observations additionnelles

$$E_{N_{i+1}}(\text{Gain}) = N_{i+1} \int_{\hat{x}^*}^{x^{i,h}} (x^{i,h}-x)\,[1-\hat{\Phi}_i(x)]^{N_{i+1}-1}\,\hat{\delta}_i(x)dx \quad \text{où } \hat{x}^* = \hat{a}_i \ .$$

- Si l'une ou l'autre des deux conditions suivantes est remplie, arrêter l'exploration; sinon passer en 3.

a) $E_{N_{i+1}}(\text{Gain}) < \beta_i . PA(x^{i,h})$

$< \beta_i (x^{i,h} - \hat{x}^*)$

où β_i est la proportion minimale du potentiel d'amélioration définissant le seuil $\beta_i(x^{i,h} - \hat{x}^*)$ que l'espérance du gain devrait dépasser pour suggérer la poursuite de l'exploration. L'indice i associé au paramètre β signifie qu'une autre proportion limite peut être choisie à chaque itération. Le choix d'une valeur pour le paramètre β_i est arbitraire. Ce choix est simplement un moyen de rendre explicite la signification de la décision résultant de l'examen du rapport entre l'espérance du gain $E_{N_{i+1}}(\text{Gain})$ et le gain maximum possible $PA(x^{i,h})$.

b) $E_{N_{i+1}}(\text{Gain}).C_o < C_S.N_{i+1} + C_E$

où C_E représente le coût correspondant à l'estimation des trois paramètres de la loi Weibull, estimation qui suivrait la génération des N_{i+1} optima locaux supplémentaires. C_O et C_S ont été définis plus haut.

---

1. PPESE: plus petit nombre entier supérieur ou égal à...

3. Engendrer N_{i+1} optima locaux supplémentaires, ensuite faire $N_{i+1}=N_{i+1}+N_i$, puis poser $i=i+1$ (N_i représente donc le nombre d'optima locaux engendrés jusqu'ici) et retourner en 2.

5. CONCLUSION

Les techniques de branch and bound s'avèrent souvent inefficaces pour résoudre certains problèmes d'optimisation combinatoire de grande taille. Les procédures heuristiques qui sont indispensables pour résoudre ces problèmes n'offrent cependant aucune garantie quant à la qualité de leurs solutions.

Pour évaluer cette qualité, notre objectif a été l'estimation de l'optimum global à partir de l'information contenue dans les optima locaux engendrés par les techniques de hill-climbing. Dans une seconde étape, nous avons formulé une règle itérative ayant pour objet le choix "économique" de la taille des suites d'optima locaux qu'une telle procédure de hill-climbing doit engendrer successivement jusqu'à ce que certaines conditions d'arrêt soient rencontrées.

L'approche que nous avons présentée se fonde sur la théorie de l'inférence statistique et sur la théorie des valeurs extrêmes; elle illustre combien le recoupement de la statistique théorique et de l'optimisation mathématique peut être fructueux.

REMERCIEMENTS

Nous remercions le Professeur H. Harter de nous avoir fourni le code de son programme d'estimation des paramètres "Weibull".

BIBLIOGRAPHIE

[1] Fisher, R. and Tippett, L. (1928). Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. Proc. Camb. Phil. Soc. 24, 180-191

[2] Garey, M.R. and Johnson, D.S. (1979). Computers and Intractability. Freeman.

[3] Gibbons, J.D. (1971). Non Parametric Statistical Inference. McGraw-Hill.

[4] Golden, B.L. (1977). A Statistical Approach to the TSP. Networks 7, 209-225.

[5] Golden, B.L. and Alt, F.B. (1979). Interval Estimation of a Global Optimum for Large Combinatorial Problems. Naval Research Logistics Quarterly, Vol. 26, No. 1,69-77.

[6] Gumbel, E. (1958). Statistics of Extremes. Columbia University Press.

[7] Harter, H. and Moore, A. (1965). Maximun Likelihood Estimation of the Parameters of the Gamma and Weibull Populations from Complete and from Censored Samples. Technometrics 7, 639-643.

[8] Knuth, D.E. (1969). The Art of Computer Programming - Seminumerical Algorithms. Vol. 2. Addison-Wesley, 121-125.

[9] Lardinois, C. (1980). Problèmes d'optimisation combinatoire de grande taille et inférence statistique. Application à la planification urbaine et régionale. Thèse de Ph. D. Faculté de l'aménagement. Université de Montréal. En rédaction.

[10] Los, M. (1980). Efficient Implementation of Heuristic Algorithms for the Quadratic Assignment Problem. A paraître dans The Journal of Regional Science.

[11] Mann, N., Schafer, R. and Singpurwalla, N. (1974). Methods for Statistical Analysis of Reliability and Life Data. John Wiley and Sons.

[12] McRoberts, K. (1971). A Search Model for Evaluating Combinatorially Explosive Problems. Operations Research 19, 1331-1349.

[13] Reiter, S. et Sherman, G. (1965). Discrete Optimizing. J. Soc. Indust. Appl. Math., Vol. 13, No. 3, 864-889.

14 Weibull, W. (1951). A Statistical Distribution Function of Wide Applicability. J. Appl. Mech. 18, 293-297.

15 Zanakis, S. (1977). Computational Experience with Some Nonlinear Optimization Algorithms in Deriving Maximum Likelihood Estimates for the Three-Parameter Weibull Distribution. TIMS Studies in the Management Sc. 7, 63-77.